Which Factors Impact the Housing Market the Most?

2950 Final Project

Constantin Miranda (cym8), Daniel O'Rourke (dpo29), Anthony Cicileo (anc52)

May 14th, 2019

Generally Accessible Portion:

Overview:

As college students, we wanted to explore a topic relevant to our daily lives. On the consumer side of the market, Millennials and Generation Z are increasingly turning to the rental economy for their daily living needs, particularly housing. Taking out a mortgage now rarely appeals to a generation that values freedom over stability.

On the business side of the market, investors are increasingly using Real Estate Investment Trusts (REITs) to diversify their portfolios. REITs mimic mutual funds in that they allow individual investors to own parts of much larger real estate empires and their accompanying income.

This study examines the housing markets of the five major regions of the United States. Within each region, we selected three cities with populations ranging from small to large. For each of these cities, we developed models to determine the impact of four factors on those cities' rental index prices. These four factors include the city's population, unemployment rate, crime rate, and per capita personal income.

Both of these sides will one day impact the three of us. We will each likely be moving to new cities and renting new apartments. We will also likely be building our retirement savings and investing in financial instruments like REITs. For this reason, we wanted to better understand the impact of these four ever-fluctuating factors.

Regions/Cities Chosen, Reasoning, & Summary Statistics

To create a rounded dataset that we could analyze, we decided to take a single state from major regions in the US. From the Northeast, Midwest, Northwest, Southwest, and Southeast, we took New York, Ohio, Washington, Arizona, and North Carolina, respectively. These states have varying population, crime, income, and unemployment rates, which is why we wanted to compare these factors to the rental price index to see how each factor affects the price. We also used three cities within each state with varying sizes (one small, one medium, and one large). These sizes were determined relatively by the city population in each state. This was achieved by finding the mean population of the cities in our population dataset, and calculating which cities qualified as small, medium and large. We used half a standard deviation as the measure that separates these categories.

The statistics in **Figure 1** show the means of population, rental index, violent crime, property crime, unemployment, and income means. It also shows the the first standard deviation of the population. The population mean was higher for larger cities, which was to be expected. The rental index is higher in Washington cities compared to cities of similar sizes in other states. The violent crime mean is highest in New York cities compared to cities of similar sizes in other states. Buffalo, NY has an especially high violent crime rate of 1252.57. The property crime mean is similar in all states and cities. The unemployment mean is consistent throughout, except it is slightly higher in all cities in North Carolina. The income mean is lower for Arizona and higher in Washington when comparing cities of the same size.

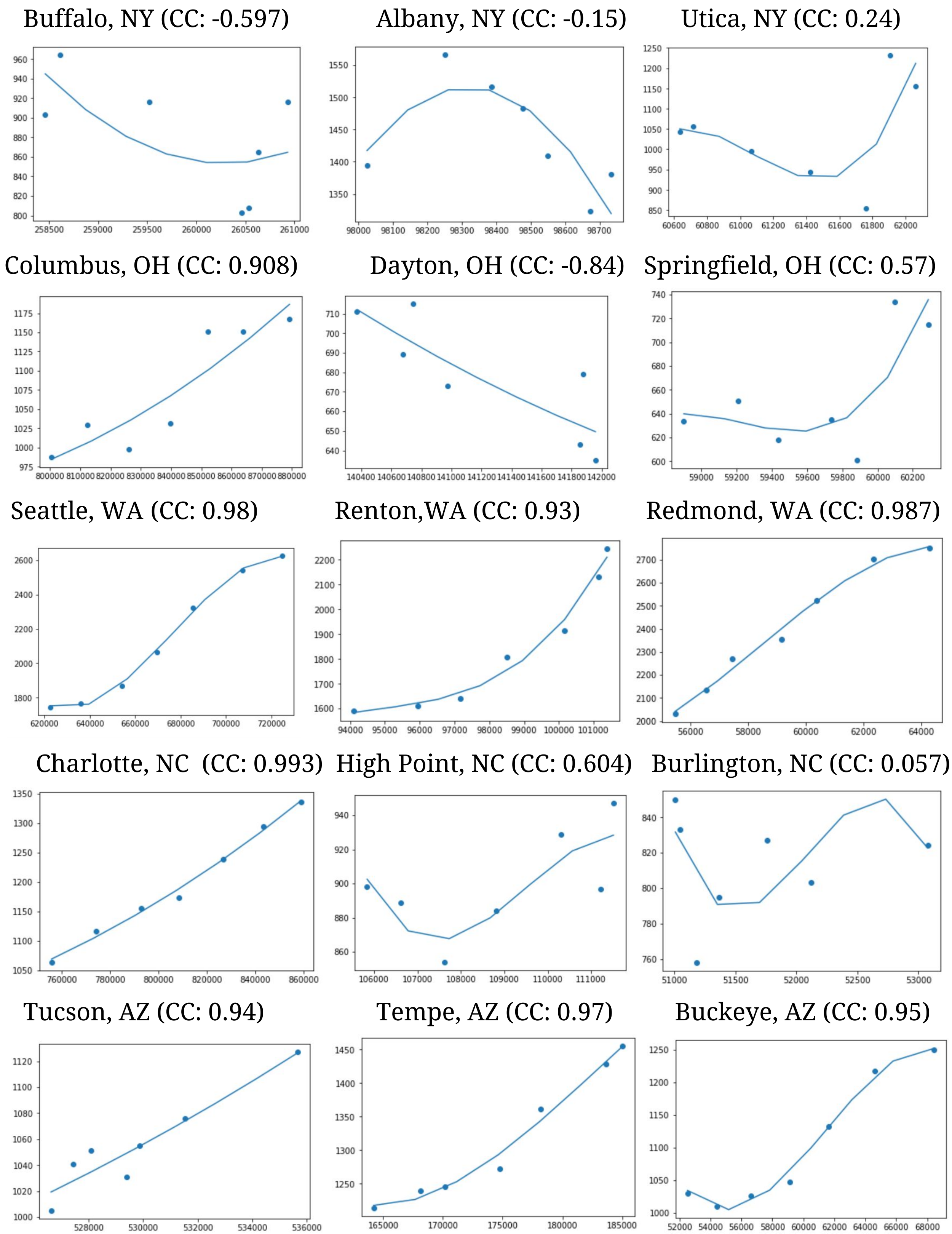
Figure 1: Summary Statistics

City	State	Size	Region	Population mean	Population Std. Dev.	Rental Index Mean	Violent Crime Mean	Property Crime Mean	Unemployment Mean	Income Mean
Buffalo	New York	Large	Northeast	259879.14	939.33	882.14	1252.57	5055.15	6.28	$44,021
Albany	New York	Medium	Northeast	98442.57	228.34	1439	845.72	4304.37	5.28	$50,726
Utica	New York	Small	Northeast	61366.85	532.04	1039.71	626.9	4073.6	5.85	$38,959
Columbus	Ohio	Large	Midwest	839061.57	26130.78	1073.42	604.92	5055.15	5.28	$45,694
Dayton	Ohio	Medium	Midwest	141207.85	619.2	677.85	918.12	5706.67	6.28	$41,827
Springfeild	Ohio	Small	Midwest	59651	458.7	655.42	680.82	7363.92	5.85	$36,330
Seattle	Washin gton	Large	Northwest	671395.71	34374.88	2133.57	594.72	4972.05	5.28	$60,468
Renton	Washin gton	Medium	Northwest	98342.14	2550.98	1848.42	279.05	4972.05	5	$71,692
Redmond	Washin gton	Small	Northwest	59376.14	2957.47	2395.14	68.7	2646.45	5	$71,692
Tucson	Arizona	Large	Southwest	529802.42	2840.66	1055.14	677.62	5911.6	5.85	$38,172
Tempe	Arizona	Medium	Southwest	174880.85	7298.98	1316.71	493.07	4892.35	5.85	$41,404
Buckeye	Arizona	Small	Southwest	59626.85	5265.29	1101.71	81.14	2393.37	5.85	$41,404
Charlotte	North Carolina	Large	Southeast	808640.42	34439.89	1196.85	613.02	3825.27	7	$45,198
High Point	North Carolina	Medium	Southeast	108843.14	2084.95	899.71	518	4166.1	6.85	$38,719
Burlington	North Carolina	Small	Southeast	51652.57	691.13	812.85	743.97	5451.1	6.71	$35,704

Models Built & Findings

(CC = Correlation Coefficient)

Figure 2: Regression: Population - Rental Index



Regression: Population - Rental Index Findings & Analysis:

Charlotte, NC:

Charlotte's housing market stands out from the rest of the markets that we analyzed. Charlotte's population has consistently risen over the past seven years and, from our regression analysis, home rental prices have steadily followed. Home rental prices are heavily related to population in Charlotte, with a correlation of 0.9933.

Dayton, OH:

Dayton's housing market stands out for another reason. Dayton's population has declined steadily over the past seven years, but their home rental prices have only increased. Dayton's population and home rental prices are negatively related, with a correlation of -0.8403. One possible explanation for this phenomenon is a recent boom in Dayton housing investment, which has led to properties being bought up for redevelopment, thus constricting the supply of Dayton rentals (Source).

General Findings:

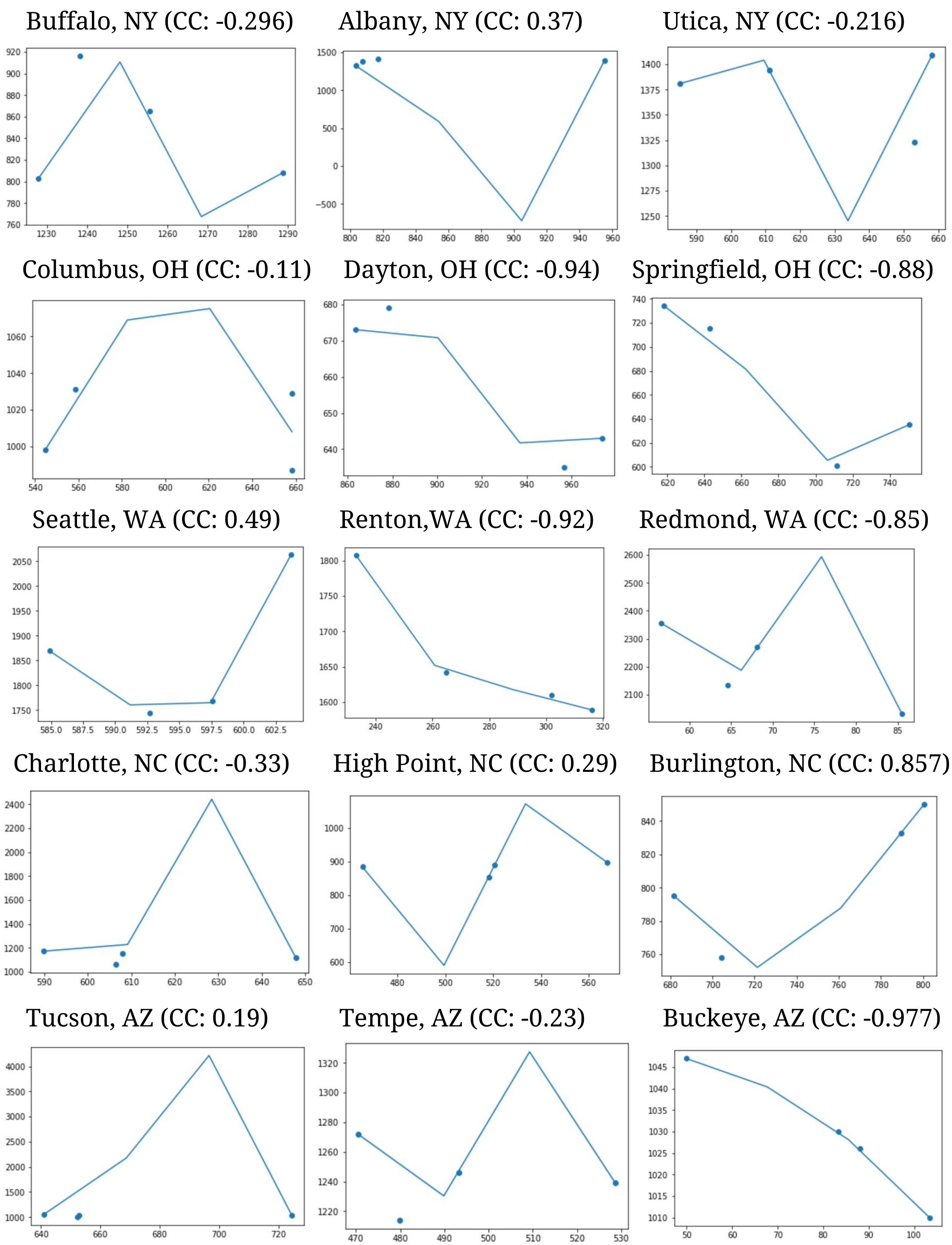
The cities in **Figure 2** are ordered by region (top to bottom) and by size (largest on the left, smallest on the right). The large cities typically exhibit the highest correlation between population and home rental price. This makes sense since large cities typically have the most market flexibility in terms of availability of supply and demand. This also makes large cities excellent for sustained returns on investment.

Medium cities are slightly more volatile, with population and home rental prices not always well correlated. As seen with Dayton, a variety of factors influence housing markets and the markets' response can be unpredictable.

The smallest cities show a flattening of the correlation line once they approach larger populations. This demonstrates that smaller cities can experience booms in populations and thus in housing prices, making them good growth investment opportunities. But, that growth can just as likely collapse once demand for housing is met.

Overall, the population is typically correlated with home rental prices across the three city sizes and across the five regions.

Figure 3: Regression: Violent Crime - Rental Index



Regression: Violent Crime - Rental Index Findings & Analysis

Buckeye, AZ:

Buckeye's violent crime vs rental price index graph (**Figure 3**) is the most negatively correlated with a correlation coefficient of -0.9770. This means that as the rental index decreases, the violent crime rate in Buckeye increases. This correlation is logical because as violent crimes in an area increase, you would expect housing prices in that area to decrease.

Burlington, NC:

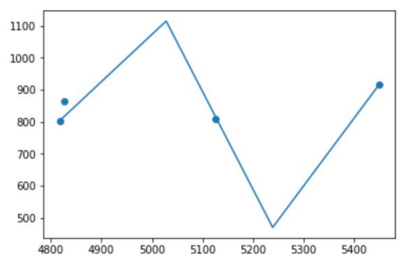
Burlington has the highest positive correlation between violent crime and rental price index. With a correlation coefficient of 0.8568, the graph in **Figure 3** shows this positive relationship, which means that as the rental index price increased, the amount of violent crimes increased. While it is not certain, this could be attributed to people desperate to pay for their housing because as market demand for rental houses increases so does price.

General Findings:

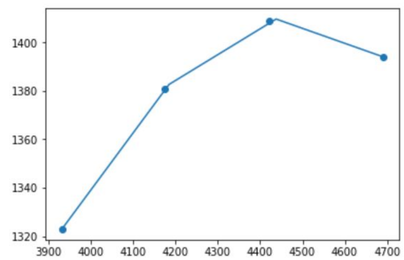
The cities in **Figure 3** are ordered by region (top to bottom) and by size (largest on the left, smallest on the right). There was no specific similar correlations among any of the states or any of the city sizes. Being that there are no general trends in the correlations between violent crime and rental index, our data shows that there is not widespread correlation between these two variables.

Figure 4: Regression: Property Crime - Rental Index

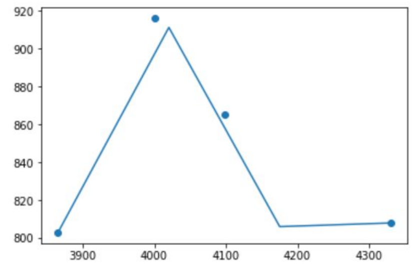
Buffalo, NY (CC: 0.64)



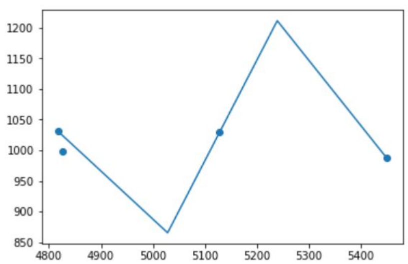
Albany, NY (CC: 0.81)



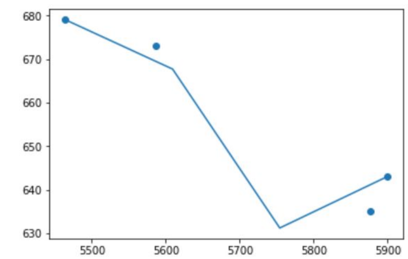
Utica, NY (CC: 0.54)



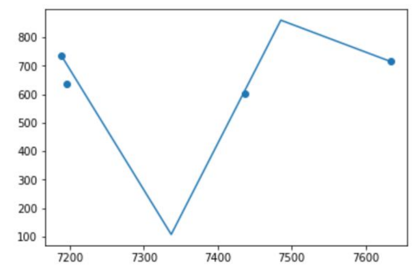
Columbus, OH (CC: -0.5)



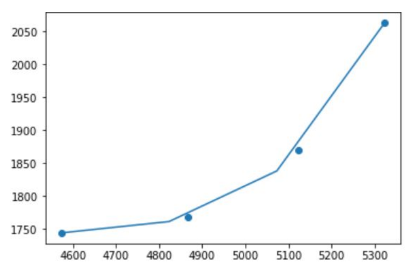
Dayton, OH (CC: -0.97)



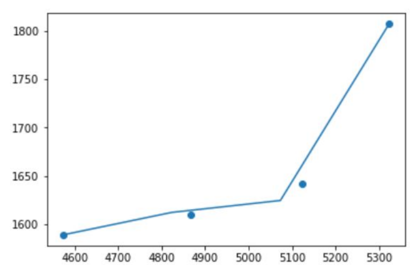
Springfield, OH (CC: 0.043)



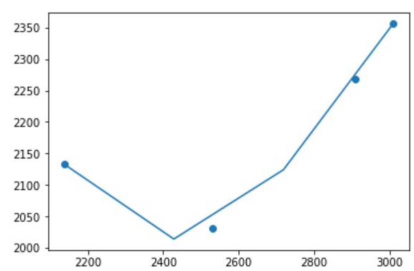
Seattle, WA (CC: 0.9)



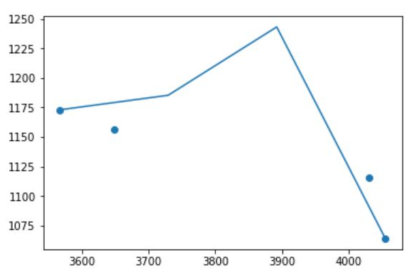
Renton,WA (CC: 0.85)



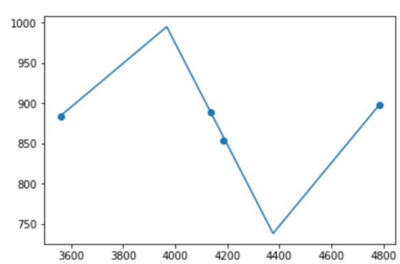
Redmond, WA (CC: 0.75)



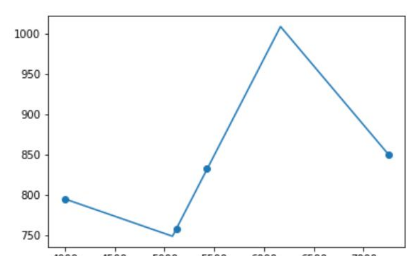
Charlotte, NC (CC: -0.91)



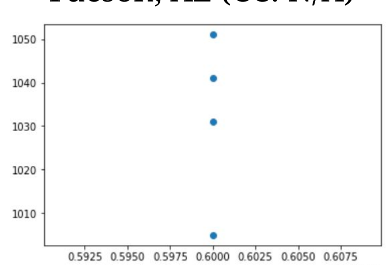
High Point, NC (CC: 0.277)



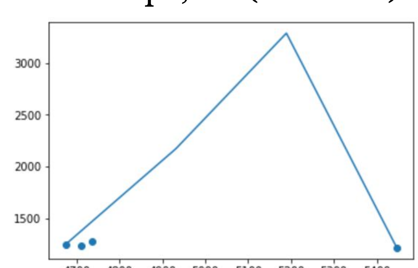
Burlington, NY (CC: 0.66)



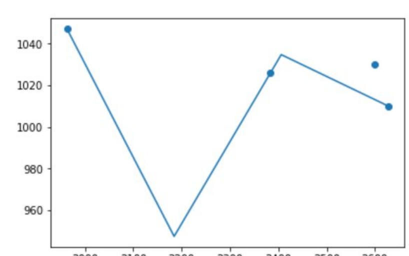
Tucson, AZ (CC: N/A)



Tempe, AZ (CC: -0.77)



Buckeye, AZ (CC: -0.855)



Regression: Property Crime - Rental Index Findings & Analysis

Tucson, AZ:

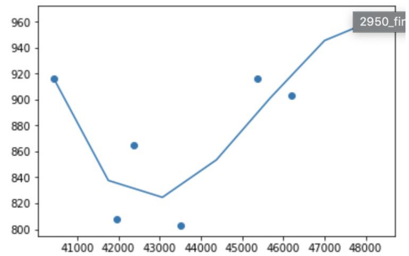
We did not have the proper data to generate a graph for Tucson, AZ; however, we did not want to eliminate property crime data as a whole because we had the data for all other states and cities. Also, property crime and rental index are inherently related because crime specific to homes will have an affect on the price of the home.

General Findings:

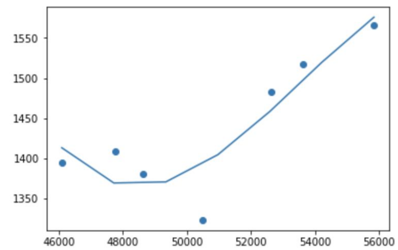
The cities in **Figure 4** are ordered by region (top to bottom) and by size (largest on the left, smallest on the right). Washington State as a whole captured our eyes while analyzing **Figure 4**. The three cities in Washington we looked at all have a positive correlation between property crime and rental index. This shows, based on our data, that as the renal index increases in Washington, property crime increases. Seattle has the highest correlation coefficient of 0.9087, followed by Renton with a coefficient of 0.8536, followed by Redmond with a coefficient of 0.7513.

Figure 5: Regression: Per Capita Personal Income - Rental Index

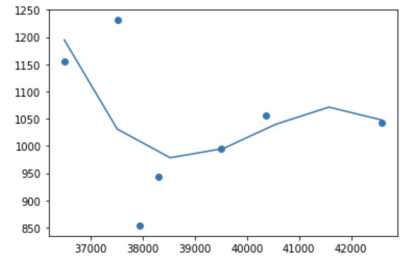
Buffalo, NY (CC: 0.55)



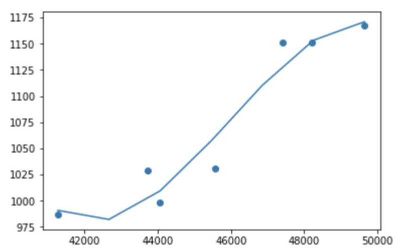
Albany, NY (CC: 0.786)



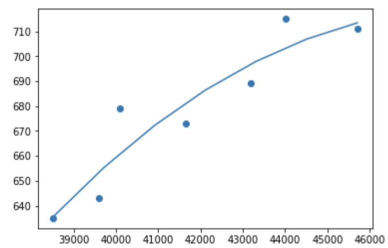
Utica, NY (CC: -0.19)



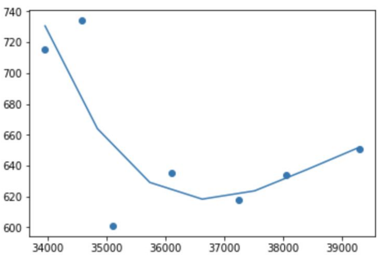
Columbus, OH (CC: 0.93)



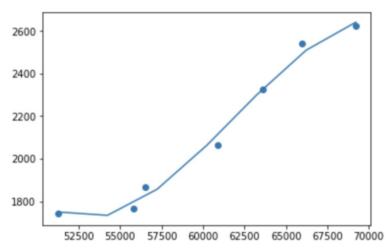
Dayton, OH (CC: 0.92)



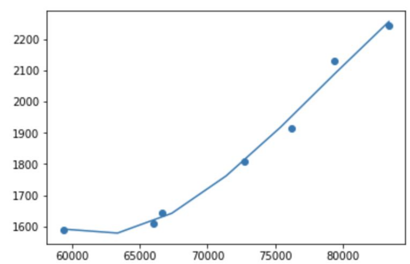
Springfield, OH (CC: -0.499)



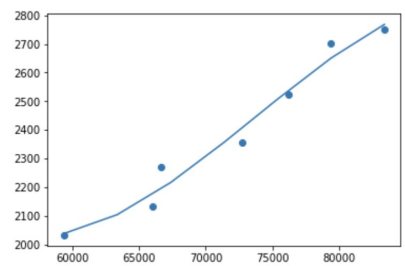
Seattle, WA (CC: 0.94)



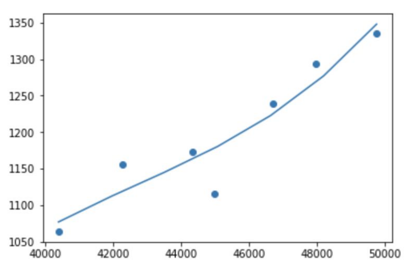
Renton,WA (CC: 0.96)



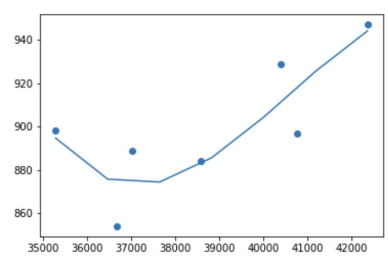
Redmond, WA (CC: 0.98)



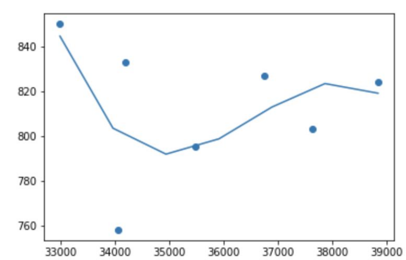
Charlotte, NC (CC: 0.93)



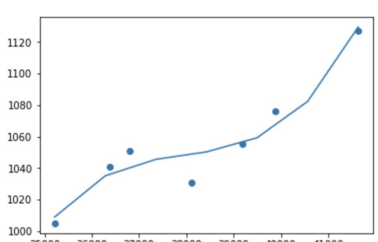
High Point, NC (CC: 0.72)



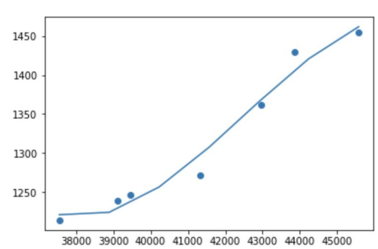
Burlington, NC (CC: -0.017)



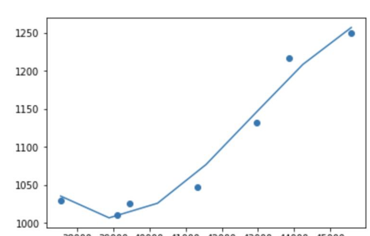
Tucson, AZ (CC: 0.9)



Tempe, AZ (CC: 0.97)



Buckeye, AZ (CC: 0.93)



Regression: Per Capita Personal Income - Rental Index Findings & Analysis:

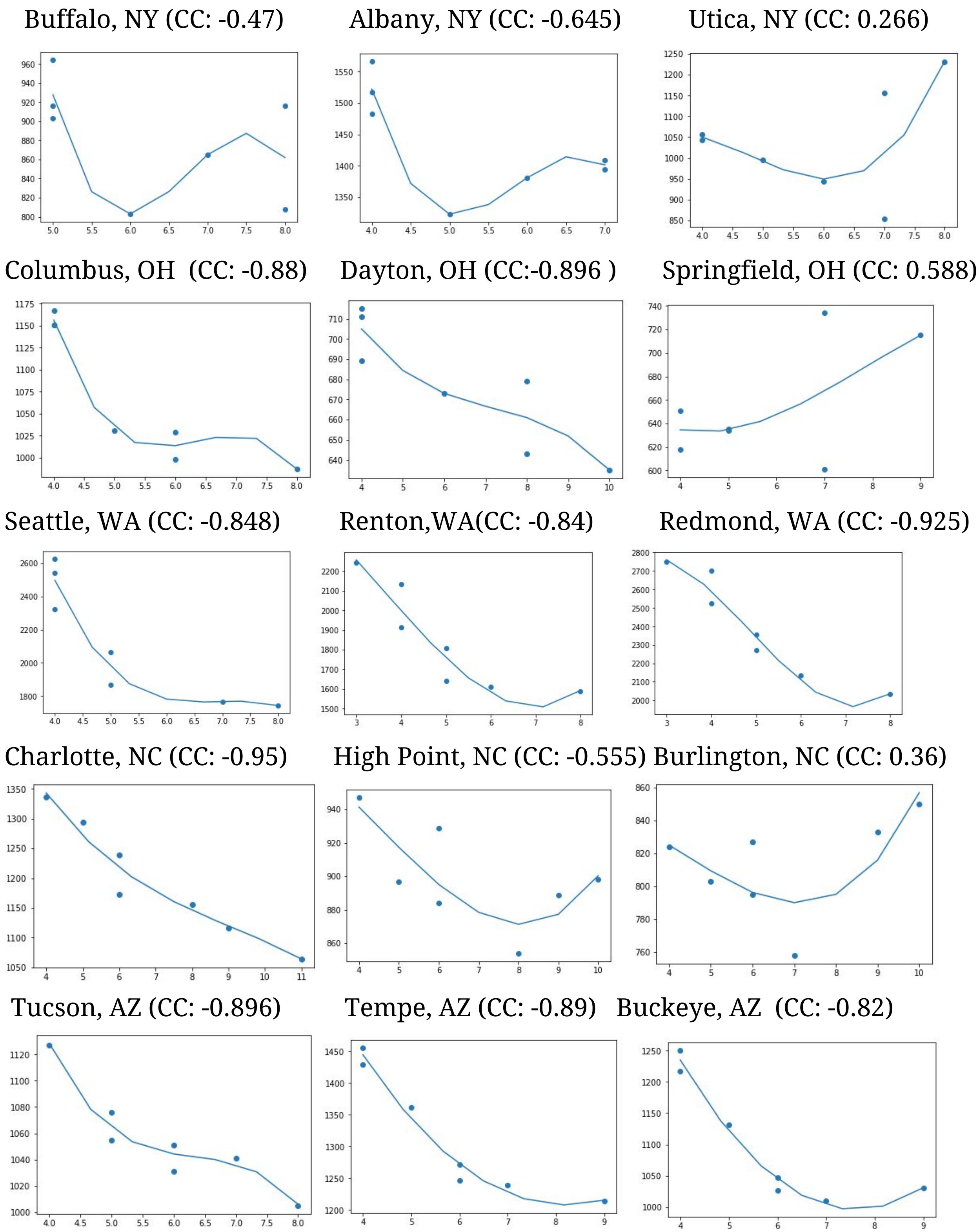
Redmond, WA

In Redmond, per capita personal income is highly correlated (0.9799) with the rental index. Redmond's rental index prices are also higher than most other cities. This makes sense since Redmond is home to several very successful companies, including Microsoft and Nintendo. Successful companies bring high salaries with them and this has likely impacted Redmond's housing market, consistently driving up prices.

General Findings:

The cities in **Figure 5** are ordered by region (top to bottom) and by size (largest on the left, smallest on the right). Overall, per capita personal income is heavily correlated with Zillow's rental index. Per capita personal income in all of these cities increased over the time period and this higher monthly income has likely fueled higher monthly rent prices. There does not seem to be a distinguishing factor between small, medium, and large cities based on these regression models. However, income across the United States rises at a pretty consistent rate, so the heavy correlation and lack of true outliers makes sense.

Figure 6: Regression: Unemployment - Rental Index



Regression: Unemployment - Rental Index Findings & Analysis

Charlotte, NC:

Charlotte's unemployment rate had the most significant correlation, of all the cities, with its mean rental index over time (-0.95). This exemplifies a phenomenon of the average rental price decreasing, as the unemployment rate in the city has increased over time. Almost all large and medium cities in this dataset have exhibited a similar, negative correlation coefficient close to -1.0. As the North Carolina cities become smaller, the shape of the line becomes less linear and more quadratic in shape.

General Findings:

The cities in **Figure 6** are ordered by region (top to bottom) and by size (largest on the left, smallest on the right). Overall, the medium sized cities within each state all have a negative correlation between unemployment and rental index. The states North Carolina and Arizona also have a negative correlation between unemployment and rental index in all cities. This means that as the rental index decreases, unemployment decreases.

<u>Scientifically Rigorous Section:</u>

Data Sources: How We Accessed Data

1. **Quandl** - Our household income, unemployment, and rental index data was accessed via Quandl's API and accompanying python package. Quandl makes many sources of public data easily accessible for developers.

   Our notebook installs the Quandl package using `!pip install quandl`. We then import the package and set our API key using the `api_key` parameter in the package configuration. To query the API, we use `quandl.get(key)`with the key corresponding to the 'table key'.

2. **CSV** - Our population data is from the US Census and was downloaded in .csv format. We used `pandas.read_csv()` to read the .csv into a Pandas Dataframe.

3. **Web Scraping** - It was difficult to find an aggregated dataset for crime statistics by city, but luckily we found the FBI's Uniform Crime Reporting Statistics Tool. We used BeautifulSoup to read and scrape both violent crime and property crime statistics for each city examined in this study.

Data Sources: How we Cleaned the Data Up/Spliced It

1. **Quandl: Population Data** - The population data was already clean and in the proper order, so we just filtered for the years that we were interested in (2011 through 2017). We then return the final array of population data.

2. **Quandl: Zillow Rental Index Data** - We wanted the Zillow data to line up with the Population Data, so we extracted each of the rows corresponding to July in the years 2011 through 2018. We appended these to a final array and return that array.

3. **Crime Data Scraping** - Our function uses `soup.find_all("tr")` to locate each of the table rows. Then, within each table row, we use `soup.find()` to locate each of the table data tags with the specified header. The header is determined by whether we are querying violent or property crime data. Once the data is located, we extract it using `.contents` and used a regular expression to extract the number strings. To finish the formatting, we stripped the strings, removed commas, cast

them to floats, and zipped them into a dictionary with an array of years ranging from 1985 to 2014.

Regression Analysis

To conduct our regression analyses, we first converted all of our lists into Numpy arrays. This allowed us to feed our data into SciKit Learn's `LinearRegression` function with the fit set to the minimum and maximum of the dataset.

We used `make_polynomial()` to convert our input Numpy array into a Pandas dataframe. It also returns the linear, squared, and cubed values from the dataframe for the regression model to use.

To set our intervals, we used SciKit Learn's `linspace()` function with the dataset minimums and maximums as parameters. We set the interval value to the length of the respective dataframe.

Finally, we used SciKit Learn's `predict()` function to create a set of predictions. We plotted the intervals and predictions on a PyPlot scatter plot of the points from the dataframes. This resulted in the regression graphs for each city based on the two variables fed in.

We also grouped cities into small, medium, and large subsections and combined their datasets for the Zillow Rental Index and the population. We then performed regression analysis on these combined datasets. We also combined the cities in each state into datasets, and performed regression analysis on each of these five "region" datasets as well.

For each of the groupings in both the state-by-state analysis and the state-size analysis, we calculated the coefficient of determination, which demonstrates to what degree the model built can predict rental price as a function of population.

Summary Statistics

Figure 7: Coefficient of Determination (CoD) Comparison Across Cities

City	State	CoD: Income	CoD: Unemployment	CoD: Population	CoD: Crime - Violent	CoD: Crime - Property
Buffalo	New York	0.3038	0.2236	0.3562	0.0877	0.4120
Albany	New York	0.6173	0.4163	0.1899	0.1383	0.6612
Utica	New York	0.0368	0.0709	0.0581	0.0464	0.2920
Columbus	Ohio	0.8740	0.7781	0.8248	0.0128	0.2511
Dayton	Ohio	0.8506	0.8021	0.7063	0.8786	0.9490
Springfield	Ohio	0.2488	0.3452	0.3259	0.7810	0.0019
Seattle	Washington	0.9408	0.7199	0.9661	0.2425	0.8259
Renton	Washington	0.9157	0.7102	0.8679	0.8513	0.7288
Redmond	Washington	0.9603	0.8564	0.9733	0.7276	0.5646
Tucson	Arizona	0.8139	0.8036	0.8886	0.0367	No Data
Tempe	Arizona	0.9400	0.8034	0.9462	0.0516	0.5987
Buckeye	Arizona	0.8654	0.6774	0.8975	0.9546	0.7315
Charlotte	North Carolina	0.8639	0.9067	0.9867	0.1083	0.8369
High Point	North Carolina	0.5180	0.3082	0.3651	0.0847	0.0768
Burlington	North Carolina	0.0002	0.1323	0.0032	0.7342	0.4400

Conclusion

From the models we built and the statistics we extracted, we can determine how particular variables impact various housing markets. Washington State's housing markets respond particularly well to income. Rental prices typically follow the trajectory of per capita personal income.

Seattle, Redmond, Tempe, and Charlotte are all markets sensitive to population. Charlotte is also sensitive to unemployment, with rental prices following unemployment trends. Increasing populations in these cities has led to increases in rental prices. Utica's housing market is interesting particularly because it does not seem to track any of these variables very closely.

Overall, population and income seem to have the most consistent impact on cities' rental prices. It is interesting to see the pocket of correlation in "high-tech" cities like Seattle and Redmond between income and rental prices. This is indicative of a macro trend in areas as wealthy as these. As expected, many cities have a low coefficient of determination when it comes to crime, both property and violent. There is a moderate and consistent correlation between the rental index and unemployment in most states. Finally, we noticed that small cities in particular are not very receptive to changes across these four variables (income, unemployment, crime, and population).